1-21-2015

Simulation-Based Inference in Statistics Education: Exciting Progress and Future Directions

Nathan L. Tintle
Dordt College, nathan.tintle@dordt.edu

Simulation-Based Inference in Statistics Education: Exciting Progress and Future Directions

Abstract

Statistics Views asked Dr Tintle to explain more about simulation-based inference in statistics education and some of the exciting progress that has been made to date and future directions within the undergraduate curriculum.

Keywords

simulation-based curricula, inference, high school students, college students, courses of study

Disciplines

Statistics and Probability

Simulation-based inference in statistics education: Exciting progress and future directions

Computational advances in the last 20 years have revolutionized the practice of statistics through the pervasive use of resampling methods, Bayesian inference, the E-M algorithm and, more recently, data mining techniques. While these methods are now commonly part of upper-level undergraduate and graduate training programs in statistics and data science, it is only recently they have begun to have been given serious consideration in the K-12 and early undergraduate statistics curriculum. Following the adoption of the Guidelines for Assessment and Instruction in Statistics Education by the American Statistical Association in 2005 [1] which focus mainly on how to teach statistics, it was the right time to look carefully at what we teach. While not a new idea, George Cobb's seminal paper in 2007 [2] challenged statistics educators to think about the pedagogical value of using simulation-based inference in introductory statistics education to facilitate student's conceptual understanding of the key pillars of statistical inference: both its logic and its scope.

In essence, simulation-based inference methods can help instructors focus on the key concepts in statistical inference instead of getting lost in a sea of mathematical algorithms and computations. For example, using this approach, students can conduct a significance test of a single proportion on day one of the course knowing no formal probability theory, instead using a simple binomial simulation (e.g. coin flipping or spinning a spinner) to see how unlikely their observed result is. Later in the course students can do a permutation test (initially modelled by shuffling cards) to simulate the re-randomization of subjects to treatment groups and evaluate the statistical significance of an observed treatment effect, without needing to first discuss a *t*-distribution and its degrees of freedom.

Since 2007, there has been increasing momentum around the use of simulation-based inference methods in teaching introductory statistics. Here are a few of the most notable events:

- The core standards for K-12 education [3] have an increased focus on statistics, and suggest the use of simulation and randomization in their implementation
- Countless conference sessions, panels and workshops have been offered within the last few years on the topic of randomization and simulation: with sessions often standing room only
- A handful of textbooks have been published which integrate randomization and simulation methods in the teaching of introductory statistics [4]
- The creation of an online community/listserv for discussion of the approach [5]

Recently, this momentum is bearing fruit. Users of these methods in their classes tend to continue using them and argue convincingly to others for their widespread use. Textbooks continue to be written and published and are growing in popularity. Even more convincingly, students are performing better on standard tests of conceptual understanding at the end of the course [6], and retaining this information longer after the course is complete [7].

Despite this momentum and progress, there are still significant questions about what the future of simulation-based inference methods will be in the undergraduate curriculum. Academia moves slowly, and the traditional, fairly mathematical, way of teaching introductory statistics persists as the norm. One of the biggest hurdles is that most introductory statistics teachers were trained as mathematicians, with little to no formal training in statistics. Furthermore, what little statistics most of these teachers did receive was in a mathematical statistics course and not a course that emphasized statistical thinking (understanding data, importance of data production, omnipresence of variability and quantification and explanation of variability [8]).

So, a course built on a foundation of simulation-based inference, which inherently emphasizes statistical thinking, is foreign and uncomfortable for many. Continued training of statistics instructors is needed so that introductory statistics courses train students to think statistically.

In addition to the need for statistics instructors to better understand and appreciate the need to teach statistical thinking, other major questions about the use of simulation-based inference methods remain. For example, how much simulation-based inference is enough? At what point should we transition new learners of statistics to traditional, asymptotic approaches? Relatedly, when learning gains are observed on key concepts (e.g. the logic of statistical inference), what is causing those gains? On the one hand simulation-based inference is conducive to a hands-on, active learning pedagogy that is widely considered a 'best-practice' in education. On the other hand, the innate simplicity of simulation-based inference may contribute to students increased ability to see the overarching statistical process of drawing conclusions from data.

Looking ahead, I am optimistic that as we continue to learn more about how, why and what students learn better in an introductory statistics course that emphasizes simulation-based inference, we will be able to leverage this better understanding throughout the undergraduate statistics curriculum, improving statistical thinking for the next-generation of statisticians, while simultaneously increasing the pipeline of individuals who can think statistically [9].

If you want to learn more about the use of simulation-based inference methods in teaching introductory statistics I encourage you to visit: www.causeweb.org/sbi to hear from a variety of statistics educators about their experiences, get access to resources, join ongoing assessment projects and join the conversation. Regardless of how you feel about simulation based inference in the introductory statistics course, the impact of computational capacity on the field of statistics is undeniable. How will you leverage this capacity to improve your student's ability to think statistically?

References

[1] http://www.amstat.org/education/gaise/

[2] https://escholarship.org/uc/item/6hb3k0nz

[3] http://www.corestandards.org/Math/

[4] https://www.causeweb.org/sbi/?p=30

[5] www.causeweb.org/sbi

[6] http://www.amstat.org/publications/jse/v19n1/tintle.pdf

[7] https://www.stat.auckland.ac.nz/~iase/serj/SERJ11%281%29_Tintle.htm

[8] http://www.amstat.org/education/gaise/

[9] http://www.amstat.org/education/curriculumguidelines.cfm